Analysis of the *Drosophila* EGFR responsive gene *CG4096*, which encodes an ADAMTS protein

Research Thesis

Presented in partial fulfillment of the requirements for graduation *with research distinction* in the undergraduate colleges of The Ohio State University

by

Sarah E. Halula

The Ohio State University

June 2012

Project Advisor: Dr. Amanda Simcox, Department of Molecular Genetics

Abstract

In *Drosophila*, a novel target of the Epidermal Growth Factor Receptor (EGFR) signaling pathway, *CG4096* was identified using microarray analysis. The gene belongs to the family of ADAMTS (A Disintegrin and Metalloproteinase with Thrombospondin motifs) proteins, which are involved in regulating components of the extracellular matrix (ECM). Tissue culture cells (S2) that stably express *Drosophila* EGFR under an inducible promoter (S2-DER II) were used to test whether induction can be demonstrated *in vitro*. Examining expression using Reverse transcription-PCR analysis of S2-DER cells showed that *CG4096* was induced, confirming the microarray data. In order to analyze the function of *CG4096*, a loss of function mutant is being generated by remobilizing a transposable element inserted in the gene. Due to conservation, analyzing the phenotype of a *CG4096 Drosophila* mutant will likely shed light on the role of *ADAMTS* genes in vertebrates. This is important because the genes are implicated in multiple disease processes.

Acknowledgements

I would like to thank the members of the Simcox Lab for assisting me in my research over the last two years (Christina Austin, Peter Halpin, and Dr.Tom Jacobsen). I would especially like to thank Sathiya Manivannan for the countless hours of mentoring, advice, help on experiments, and morale support; this thesis would not have been possible without him. I would also like to give a special thanks to Dr. Amanda Simcox for allowing me to conduct research in her lab and taking time to mold me into a research scientist. She has helped me discover my passion for molecular biology and discover the courage to present scientific ideas. I would like to thank The Ohio State University fly community and Dr. Harald Vaessin for serving on my committee. Finally thank you to the College of Arts and Sciences for the Undergraduate Research Scholarship to support my work.

Table of Contents

Introduction

The epidermal growth factor receptor (EGFR) pathway is a highly conserved cell signaling pathway that is required for cell proliferation, differentiation, migration and survival during the development of an organism— including vertebrates and *Drosophila*[1]. *Drosophila melanogaster*, serves as a model organism for studying EGFR signaling because there is only one epidermal growth factor receptor in the fruit fly compared to four epidermal growth factor receptors in vertebrates; therefore it is a simplified model for studying the EGFR pathway.

EGFR is a member of the receptor tyrosine kinase family (RTKs). As is characteristic of RTKs, EGFR, has a tyrosine amino acid side chain that upon phosphorylation—which is catalyzed by the receptor's kinase domains— causes activation of the pathway. In *Drosophila,* EGFR is activated by four ligands: *keren, spitz*, and *gurken* of the TGF-alpha family and *vein*, a neuregulin[2]. Once bound, the ligands trigger an internal signal transduction pathway that activates target genes (Figure 1). This transduction pathway is the RAS/RAF/MEK/MAPK pathway. The MAPK kinase activates ETS transcription factors that regulate target genes[3]. When EGFR signaling is received by surrounding cells, they activate or repress responsive genes leading to modification of cellular activity[4].

Genetic evidence supports the idea that a number of target genes of EGFR in *Drosophila* act in negative or positive feedback loops that alter the pathway[5]. Negative inhibitors of the EGFR pathway in *Drosophila* that have been identified include: *argos, sprouty, kekkon-1, MAPK Phosphatase 3, mae, and d-Cb* [6]. Positive activators of the EGFR pathway have also been identified and include: *vein, spitz, pointed*, and an *miRNA, miR7*[7]. The identification and

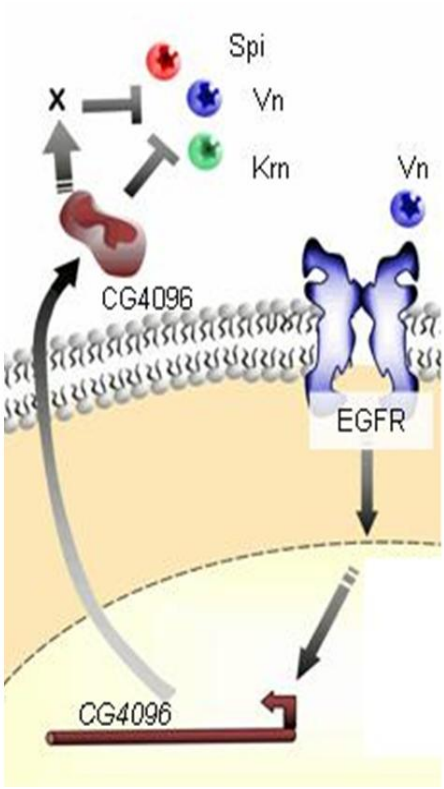
analysis of new gene targets of EGFR signaling in our recent publication adds a number of genes to these lists[8].

The identification of novel targets of the epidermal growth factor receptor pathway is a goal in the Simcox lab. A microarray analysis was done using third instar larvae wing disc RNA. The microarray analysis identified 29 previously uncharacterized candidate target genes of EGFR. One such gene was *CG4096*. Following the identification of *CG4096* as a candidate target of EGFR, *in situ* hybridization of *CG4096* was done on wild type wing discs, wing discs under a 71B-GAL4 driver to constitutively over express activated EGFR (71B>Egft^{TOP}), and wing discs under 71B-GAL4 driver to constitutively express dominant negative (decreased) EGFR signaling (71B>Egft^{DN}). The results of the *in situ* hybridization supported the findings of the microarray analysis. In wild type wing discs, tissue corresponding to future veins showed expression (Figure 2, A). Since vein formation is guided by EGFR signaling, it is indication that *CG4096* is a target in the pathway[9]. In 71B>Egft^{TOP} wing discs in which the EGFR signaling is over expressed, the expression of *CG4096* was elevated; this again is indication that *CG4086* is a target of EGFR signaling (Figure 2, B). In 71B>Egft^{DN} wing discs in which EGFR signaling is decreased, the expression pattern of *CG4096* is decreased; yet another indication that *CG4096* is a target of the pathway (Figure 2, C).

The Simcox lab also analyzed RNAi knockdown of *CG4096* (Figure 3). After the knockdown of *CG4096*, an extra vein phenotype appeared between the margin and longitudinal vein 2 on the wing (Figure 3, B). Because EGFR is induced in ectopic tissue where it is not normally expressed, it is indication that *CG4096* is part of a negative feedback loop that inhibits EGFR signaling[10].

Figure 1: A model of EGFR signaling in *Drosophila*:

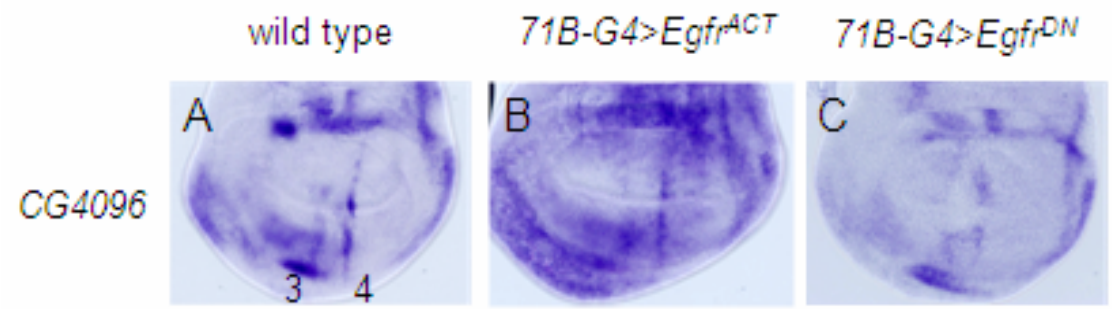
A schematic model of EGFR signaling in *Drosophila* is shown. In *Drosophila,* EGFR is activated by three zygotic ligands: *keren* and *spitz* of the TGF-α family and *vein*, a neuregulin. Because *gurken* (also of the TGF- α family only acts in embryos, it is not included in the model. Upon activation, EGFR triggers an intracellular signal transduction pathway that activates downstream target genes. The activation of *CG4096* is depicted as a downstream target of EGFR signaling. *CG4096* also functions as part of a negative feedback loop that inhibits EGFR at the level of the ligands.



(Butchar et al. 2012)

Figure 2: *In situ* hybridization of *CG4096* in Drosophila wing discs:

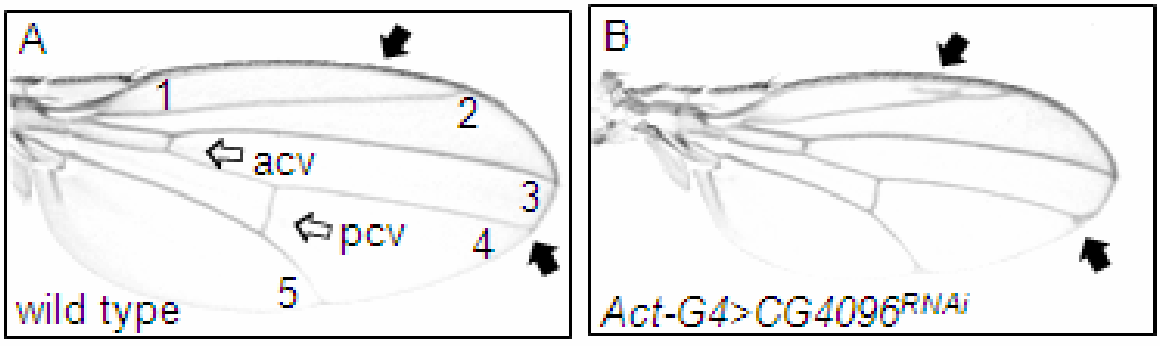
Upon *in situ hybridization* of *CG4096* in wild type wing discs (A) tissue corresponding to future veins (3 &4) showed expression shows expression. Wing discs in which a 71B-Gal4 driver was used to constitutively express activated EGFR (B) showed increased expression of *CG4096.* Wing discs in which a 71B-Gal4 driver was used to constitutively express dominant negative EGFR signaling showed decreased expression of *CG4096.* These results indicate that *CG4096* is a target of EGFR signaling in *Drosophila*.



(Butchar et al. 2012)

Figure 3: RNAi knockdown of *CG4096*:

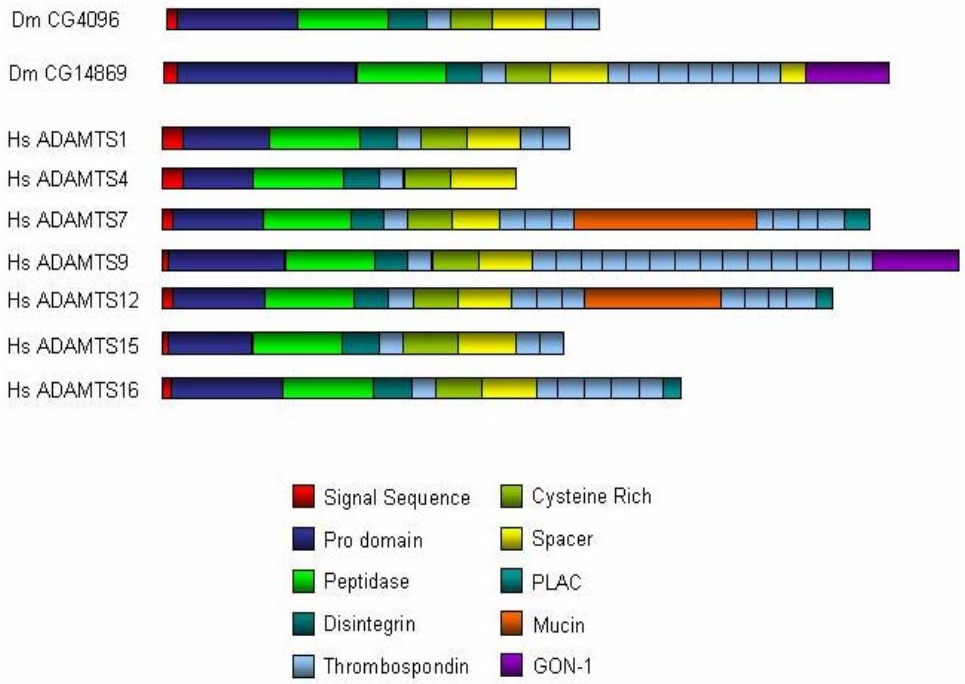
Wild type wings showed normal expression of EGFR signaling (A). After RNAi knockdown of *CG4096* (using an Actin-Gal4 driver), an extra vein phenotype was visible between the margin and longitudinal vein 2 (B). The ectopic expression of EGFR in tissue where it is not normally expressed indicates that *CG4096* is part of a negative feedback loop that acts to inhibit EGFR signaling.



(Butchar et al. 2012)

Figure 4: Domain Architecture of *Drosophila* and human ADAMTS proteins:

The domain architecture of ADAMTS proteins CG4096 and CG14869 are depicted here with 7 of the homologous human ADAMTS proteins. Characteristic metalloproteinase, disintegrin, and thrombospondin domains of a typical ADAMTS protein can be identified in the CG4096 amino acid sequence. CG4096 is most similar to ADAMTS1 based on domain structure.



(Butchar et al. 2012)

Significance & Goals

CG4096 is an ADAMTS (A Disintegrin And Metalloproteinase with ThromboSpondin motifs) gene in *Drosophila* with homolog ADAMTS genes in humans (Figure 4)[11]. There are 19 ADAMTS genes in mammals which contain specific domains: a signal peptide, a pro-domain, a metalloproteinase catalytic domain with a reprolysin-type zinc-binding motif, a disintegrin-like domain, a thrombospondin domain, a cysteine-rich domain, a spacer region of variable length, and a variable number of C-terminal thrombospondin repeats[12]. *CG4096* is most like ADAMTS1 and its subfamily (ADAMTS4, 5, 8, 15) based on domain structure[13]. Of these *CG4096* is most similar in sequence to ADAMTS15[14]. Based on sequence analysis alone *CG4096* is most like ADAMTS7 and ADAMTS12[15]. This is significant because the ADAMTS genes have been linked to several diseases in humans. By analyzing the role of *CG4096* in *Drosophila*, the role of ADAMTS genes in vertebrates may be better understood.

ADAMTS genes are secreted metalloproteases involved in modification of the extracellular matrix[16]. Substrates of the extracellular matrix that are modified by ADAMTS genes include aggrecans, brevicans, proteoglycans, and EGF ligands. Aggrecanases, which include ADAMTS-1, -4, -5, -8, and -15, cleave aggrecan, brevican, and versican[17]. Aggrecan loss (and therefore the loss of connective tissue) caused by the uncontrolled proteolysis of aggrecans by ADAMTS proteins has been implicated in arthritis[18]. Similarly, uncontrolled proteolysis of brevicans and versicans has been implicated in atherosclerosis[19]. In atherosclerosis plaque, formed by coagulation and aggregation of platelets, builds in the arteries. Over time the obstruction of blood flow by the plaque can cause the arteries to rupture resulting in a myocardial infarction commonly known as a heart attack. ADAMTS1 and ADAMTS4 play a positive role

in the formation of plaque because of their ability to cleave versican[20]. When versican is cleaved it allows vascular smooth muscle cells to migrate and contribute to the plaque forming in the arteries[21]. While there are many human diseases linked to ADAMTS protein cleavage, the most connected to this thesis is that uncontrolled modification of EGF ligands has been implicated in cancer— as described in the next section[22].

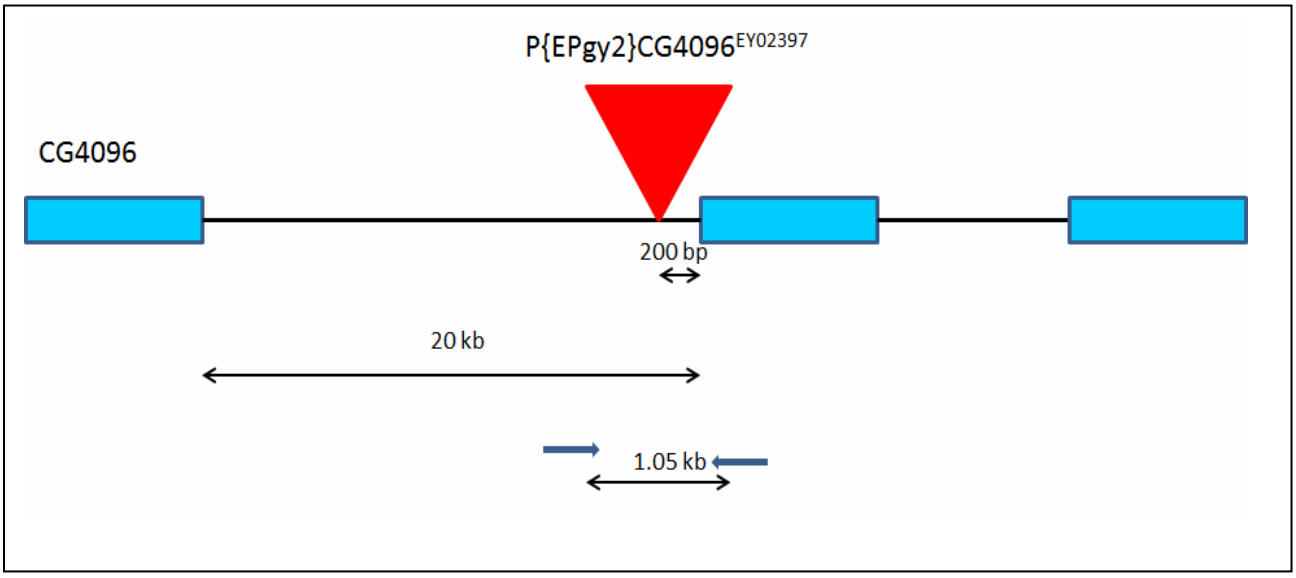
In cancer, ADAMTSs facilitate metastasis through the degradation of the extracellular matrix. Interestingly, ADAMTSs can act as tumor promoters or inhibitors. ADAMTS1, ADAMTS12 and ADAMTS15 act as tumor suppressors in prostate, colon, and breast cancer[23]. In mammals, ADAMTS1 may act as a tumor suppressor by sequestering ligands[24]. Because *CG4096* is part of a negative feedback loop, the gene may act a tumor suppressor (like ADAMTS-1) to inhibit ligands binding to EGFR[25]. Because of the implication of ADAMTS genes in human diseases, I am interested in further analyzing the functional relationship between *CG4096* and the EGFR pathway.

The first aim of my thesis is to verify *CG4096* as a transcriptional target of EGFR signaling. This was done using Schneider 2 *Drosophila melanogaster* (S2) cells that are easily cultured and used to study the expression of proteins in *Drosophila*. S2 cells are stably transfected with epidermal growth factor receptor under a metallothionein promoter. These cells are referred to as DER II cells and are induced to express the transgene using CuSO4. Comparison of the amount of RNA that corresponds to *CG4096* expression between S2 and DERII cells indicates whether a gene is a target specifically in the EGFR pathway. Identifying *CG4096* as a transcriptional target of EGFR signaling in DERII cells indicates that *CG4096* as a target of EGFR is not merely a local phenomena in the *Drosophila* wing disc.

The second aim of my project is to study the phenotypic effects of *CG4096* by producing a null mutant. A genetic approach is being used to remobilize a transposable P-element inserted into the gene in the intron before the first translational start point. (Figure 5). Mobilization of the P-element with a transposase causes both precise and imprecise excisions. An imprecise excision of the P-element removes not only the P-element but also part of the *CG4096* gene; this produces a null mutant.

Figure 5: Map of the insert site of EPgy2 in the intron of *CG4096* in *Drosophila:*

The P-element insert is located in the 20 kb intron of *CG4096* on the X chromosome 200 bp away from the first exon. Primers have been designed 500 bp from the first insertion site on either side to identify a precise or imprecise excision. A precise excision will therefore produce a 1kb band. An imprecise excision in the 1kb span will show a smaller band; however, if the imprecise excision is outside the 1kb range, no band will be present.



Material & Methods

RT-PCR:

S2 and S2-DER cells were serum starved overnight (16 hours) to slow the growth process and allow for the maximum amount of induced transcription. The cells were then treated with $CuSO_4$ (final concentration 700 μM) for 9 hours. Total RNA was extracted from the cells using a Qiagen RNeasy kit. The RNA was then used in a Reverse transcription (RT) using an Omniscript RT kit (Qiagen) to generate the complimentary cDNA. The product of the RT reaction was used as a template for PCR reactions using gene specific primers for 25 cycles. *Ornithine decarboxylase (Oda)* was used as a control in the reaction because it is expressed at equal levels in all cells. The primers used were as follows:

CG4096: 5'-CCAGGGTGCCACCTACAAG-3' (forward)

5'-GAGAATGTGCCGCGC-3' (reverse)

CG14869: 5'- GAGTCACGATGCAGTAC-3' (forward)

5'-AACCTTGACACCCTGG-3' (reverse)

Oda; 5'-GTCCTTCGGTAGAGCGACAT-3' (forward)

5'-GCACCATCTCGACTTCGTCT-3' (reverse)

Stocks:

The following gene alleles were used to generate an excision of the P-element: y^1 w^{67c23} P{EPgy2}*CG4096*^{EY02397}, w^+/y ; ; Sb(Δ2-3)/TM6, Tb^1, and PG142/FM7.These stocks were grown at 25°C.

PCR Based Genotyping:

Genomic DNA was extracted from females carrying a precise excision using a protocol from the Asano Lab. Gene specific primers were used to amplify a region of 1kb (500 bp away from the insertion site of the p-element on each side) if a precise excision had occurred. The primers were as follows:

CG4096 deletion mapping primers: 5'- GTCATCGCACTGACGGAG- 3' (Forward)

5'- TCAGCAGCATGGCAGC -3' (Reverse)

Results & Discussion

Induction of EGFR in S2-DERII cells causes the expression of *CG4096* and *CG14869*: In order to validate the previous discovery of *CG4096* as a target gene of EGFR signaling, an in vitro analysis using RT-PCR was conducted using RNA extracted from S2 and S2-DERII cells. After serum starvation, S2 and S2-DERII cells were induced with $CuSO_4$. Induction leads to the expression of target genes of the EGFR pathway. Upon reverse transcription and PCR using gene specific primers the expression of *CG4096* is only visible in S2-DERII cells induced with $CuSO_4$ (Figure 6, row 1; lane 4 (I)). The expression of *CG4096* is not visible in uninduced S2 cells, induced S2 cells, or uninduced S2-DERII cells (Figure 6: row 1; lane 1 (UI), lane 2(I), and lane 3(UI)). The expression of *CG4096* in S2-DERII cells indicates that *CG4096* is a transcriptional target of EGFR signaling. The identification of *CG4096* as a transcriptional target of EGFR signaling in S2-DERII cells is an important discovery because it shows that the expression of *CG4096* is not a tissue specific phenomenon. Therefore, the overexpression of EGFR signaling resulting in the increased expression of *CG4096*, an ADAMTS gene, may be involved in disease in vertebrates. Homologous genes to *CG4096* in humans, ADAMTS1 and ADAMTS15, have been shown to be overexpressed in cancer which makes the finding of *CG4096* as a target of EGFR signaling even more interesting.

Induction of S2 and S2-DERII cells was done in parallel with gene specific primers to express *CG14869*, another ADAMTS gene in *Drosophila*. The expression of *CG14869* is seen at a more robust level in induced S2-DERII cells (Figure 6, row 2; lane 4 (I)). Although *CG14869* expression can be seen in uninduced S2-DERII cells (Figure 6, row 2; lane 3 (UI)) (indicating

that it is also a target of another signaling pathway or that the inhibition of EGFR is leaky), the increased expression under induced EGFR signaling is indication that *CG14869* is also a transcriptional target of EGFR signaling. Further genetic analysis of *CG14869*, may shed more light on the role of *CG14869* and its relationship with EGFR.

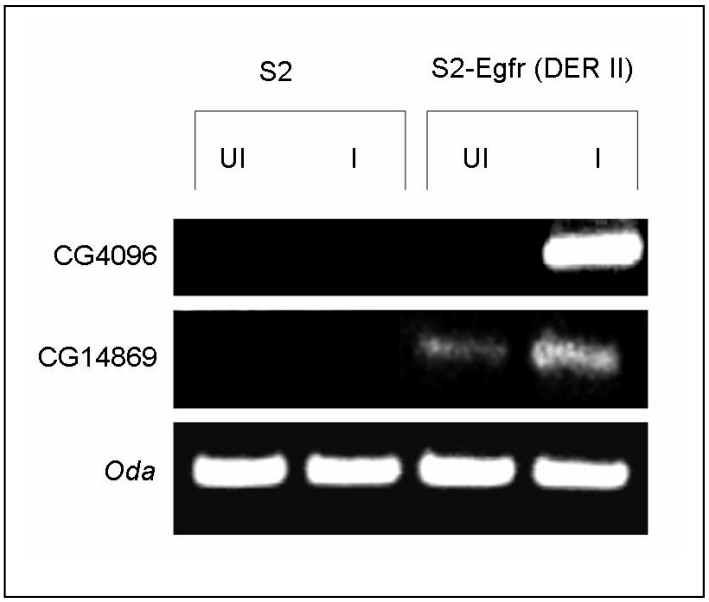
Oda, a gene expressed equally in all cells, was used as a control in the RT-PCR analysis of *CG4096* and *CG14869*. *Oda* was expressed in uninduced S2 cells, induced S2 cells, uninduced S2-DERII cells, and induced S2-DERII cells (Figure 6, row 3; lane 1 (UI), lane 2 (I), lane 3(UI), lane 4 (I)). The expression of *Oda* equally across cell types indicates that our experimental approach was valid.

Figure 6: Results of RT-PCR in S2 and S2-DERII *Drosophila* cells:

The expression of *CG4096* is present in lane 4 (I) pertaining to S2-DERII cells induced with $CuSO_4$. These results indicate that *CG4096* is a transcriptional target of EGFR. Lane 1 (UI) shows uninduced S2 cells while lane 2 (I) shows induced S2 cells. Lane 3 (UI) shows uninduced S2-DERII cells. The expression of *CG4096* is not evident in S2 cells or in uninduced S2-DERII cells.

The expression of *CG14869* is present in uninduced S2-DERII cells, but a more robust expression of *CG14869* is evident in induced S2-DERII cells. This indicates that *CG14869* is a transcriptional target of EGFR signaling, but it is also able to be expressed by alternate pathways at low levels when EGFR is not induced.

Oda , the control, is expressed equally in all four cell types. This indicates that the experimental procedure is valid.



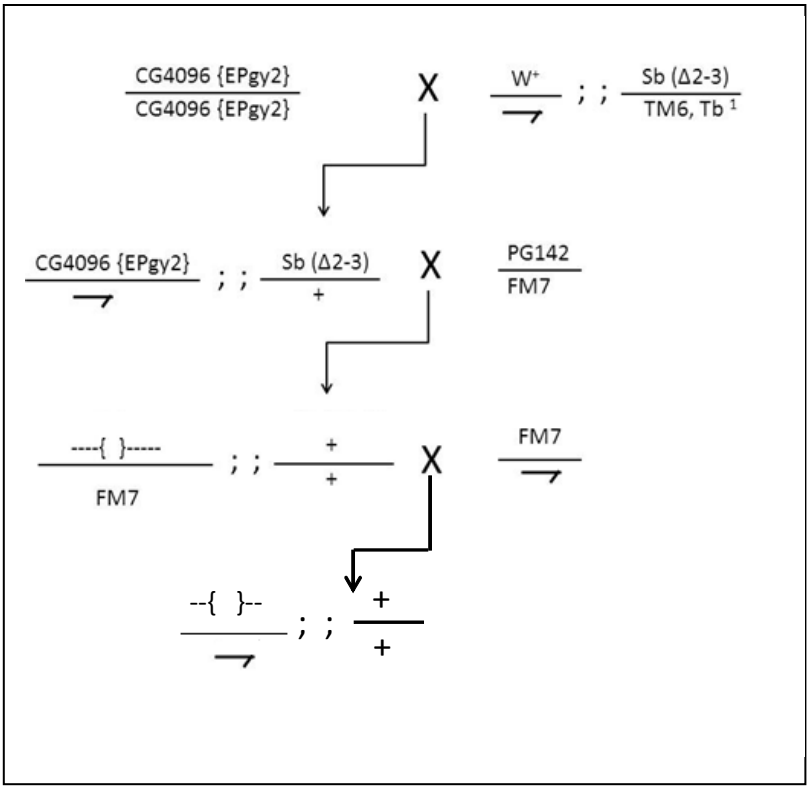
Identification of a candidate null mutant of *CG4096*:

Genetic Cross:

Virgin females from the $y^1\ w^{67c23}\ P\{EPgy2\}CG4096^{EY02397}$ stock with the P-element insert were crossed to $w^+/y\ ;\ ;\ Sb(\Delta 2\text{-}3)/TM6,\ Tb^1$ males that carried the transposase. P-elements are mobilized in the male progeny with a transposase from this cross (genotype *CG4096{EPgy2}/Y; ; Sb(Δ2-3)/TM6, Tb*[1)]. These males were crossed to virgin females from the *PG142/FM7* stock. White eyed female progeny from this cross resulting from excision of the P-element insert (genotype *–( )—/FM7*) were crossed to males from the FM7 stock. Male progeny from this cross with the excision of the P-element (genotype *–( )—/ Y*) were scored for a phenotype which could be indicative of an imprecise excision.

Figure 7: Genetic Cross to produce an imprecise excision of *Drosophila* P-element:

Females with a transposable P-element inserted into the first intron of *CG4096* (characterized by red eyes) were crossed to males carrying the transposase (characterized by short bristles). Male progeny from this cross carrying the P-element and the transposase were selected (characterized by having orange eyes and short bristles); the P-element is mobile. These males were crossed to FM7/PG142 females. Female progeny carrying the excision of the p-element (characterized by white eyes) were selected and crossed to FM7 males. The male progeny from this cross were scored for a phenotype that would indicate an imprecise excision of the P-element and therefore yield a null mutant of *CG4096*.



Phenotypic Scoring:

Male progeny with the excision of the P-element were scored for a phenotype that would indicate an imprecise excision of the insert and be potential loss of function mutants of *CG4096*. Only males were scored because *CG4096* is on the X-chromosome; phenotypic effects would be seen in hemizygous males and not heterozygous females with one normal copy of the gene. Male progeny were scored for an extra vein phenotype which would be indication of ectopic expression of EGFR signaling due to the lack of the negative feedback loop maintained by *CG4096*. A total of 134 excision events were analyzed.

Of the 134 excision events set up through a genetic screen, 134 showed no visible phenotype. There was, however, a candidate mutant that produced no males of the excision phenotype. This candidate produced 15 females with the excision, 7 FM7 females, and 16 FM7 males. This candidate, therefore, could harbor a lethal phenotype as the result of an imprecise excision of the P-element and a resulting mutation in the *CG4096* gene.

Table 1: Phenotypes of Excision Events

Number of excisions	Number of visible phenotype	Number of lethal phenotype	
134	0	1	
Distribution of progeny of lethal phenotype			
♀ --{ }-- / FM7	♂ --{ }-- / Y	♀ FM7/FM7	♂ FM7/ Y
15	0	7	16

Of the 134 excision events analyzed for a phenotype, 0 showed a visible phenotype. There was one lethal candidate that yielded 0 males with an excision of the P-element. This could suggest an imprecise excision of the P-element that caused a lethal null mutant of *CG4096*. The lethal candidate gave 15 females with an excision, 7 FM7 females, and 16 FM7 males suggesting it was a healthy and viable cross.

PCR based genotyping of a precise excision:

Gene specific primers were designed 500 bp from the original insertion site of the P-element of *CG4096* (insert located at X:5,596,544). Genomic DNA was extracted from females containing the PCR based genotyping of the genomic DNA showed a precise excision with the expected length of 1kb. This indicates that the method of mobilizing the P-element, as well as the primer design, is valid. PCR based genotyping of females from the candidate for an imprecise excision have yet to be genotyped, so the location of the imprecise excision of the P-element inserted into *CG4096* is currently unknown.

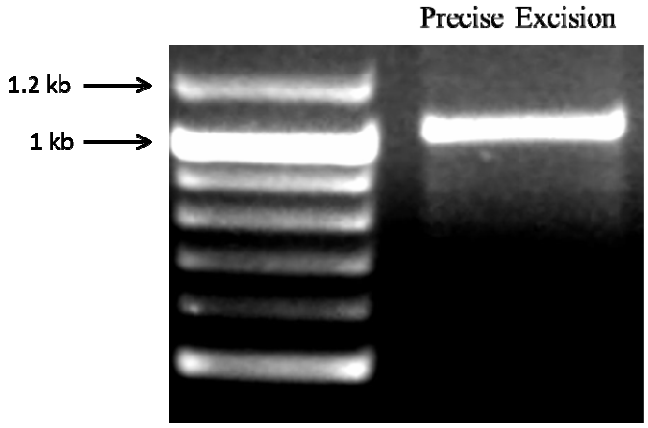


Figure 8: Precise excision of transposable P-element: The P-element has been excised from *CG4096* in the genetic screen by a transposase. PCR based genotyping of a female carrying a precise excision indicates a precise excision at a base length of 1kb.

References

[1] Shilo, Ben-Zion. Regulating the Dynamics of EGF Receptor Signaling in Space and Time. Development 2005 132 (18), 4017-4027.
[2] Shilo 2005. 4017-4027
[3] Shilo 2005. 4017-4027
[4] Shilo 2005. 4017-4027
[5] Avraham, R., and Y. Yarden, 2011 Feedback regulation of EGFR signaling: decision making by early and delayed loops. Nat Rev Mol Cell Biol 12: 104-117.
[6] Butchar, J. P., D. Cain, S. Manivannan, A. D. McCue, L. Bonanno, S. Halula, S. Truesdell, C. Austin, T.L. Jacobsen and A. Simcox, 2012. New negative feedback regulators of Egfr signaling in *Drosophila*. Genetics/2012/141093
[7] Butchar et al. 2012 Genetics.
[8] Butchar et al. 2012 Genetics.
[9] Butchar et al. 2012 Genetics.
[10] Butchar et al. 2012 Genetics.
[11] Apte, S.S., Disintegrin-like and metalloprotease (reprolysin-type) with thrombospondin type 1 motif (ADAMTS) superfamily: functions and mechanisms. J Biol Chem, 2009, 284, 31493-31497.

[12] Porter, S., I. M. Clark, L. Kevorkian and D. R. Edwards, 2005 The ADAMTS metalloproteinases. Biochem J 386: 15-27.
[13] Butchar et al. 2012 Genetics.
[14] Butchar et al. 2012 Genetics
[15] C. Nicholson, S. B. Malik, J. M. Logsdon, Jr. and E. G. Van Meir, Functional evolution of ADAMTS genes: evidence from analyses of phylogeny and gene organization BMC Evol Biol, 2005, 5, 11.
[16] S. Porter, I. M. Clark, L. Kevorkian and D. R. Edwards, Biochem J, 2005, 386, 15-27.
[17] S. Porter, I. M. Clark, L. Kevorkian and D. R. Edwards, Biochem J, 2005, 386, 15-27.
[18] Apte, S.S., J Biol Chem, 2009, 284, 31493-31497.
[19] Apte, S.S., J Biol Chem, 2009, 284, 31493-31497.
[20] Salter, R.C., T.G. Ashlin, A.P. Kwan and D.P. Ramji, 2010 ADAMTS proteases: key roles in atherosclerosis? J Mol Med (Merl) 88: 1203-1211.
[21] Salter, R.C et al. J Mol Med (Merl) 2010, 88: 1203-1211.
[22] Apte, S.S., J Biol Chem, 2009, 284, 31493-31497.
[23] Butchar et al. 2012 Genetics

[24] Liu, Y. J., Y. Xu and Q. Yu, Full-length ADAMTS-1 and the ADAMTS-1 fragments display pro- and antimetastatic activity, respectively. Oncogene,2006, 25:2452-2467.
[25] Butchar et al. 2012 Genetics